LIBRARY OF THE
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN
510.84
Il6r
no.782-787
cop. 2

The person charging this material is responsible for its return to the library from which it was withdrawn on or before the **Latest Date** stamped below.

Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.

UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

JUN 6 1973

MAY 9 RECD

L161—O-1096

Digitized by the Internet Archive
in 2013

http://archive.org/details/detailsofexperim782dori

UIUCDCS-R-76-782

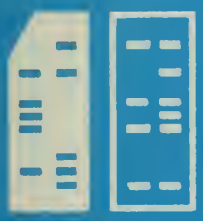
DETAILS OF AN EXPERIMENTAL VIDEOTAPE EVALUATION OF AN INTERACTIVE EXAM SYSTEM

by

Richard Doring
Lawrence R. Whitlock
Wilfred J. Hansen

the Library of the
APR 12 1977
University of Illinois

December 1976



DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN · URBANA, ILLINOIS

510.84
Il6r
no.782-787
cop.2

UIUCDCS-R-76-782

DETAILS OF AN EXPERIMENTAL VIDEOTAPE EVALUATION OF AN INTERACTIVE EXAM SYSTEM

by

Richard Doring
Lawrence R. Whitlock
Wilfred J. Hansen

December 1976

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801

This work supported in part by the National Science Foundation under grant EC41511.

TABLE OF CONTENTS

Page

1. Introduction 1

2. Environment. 2

3. Experimental Procedure 6

4. Summary of Results 9

5. Design Proverbs. 14

5.1. Instructions and preparation 16
5.2. Problem generation and display 17
5.3. Entering answers 19
5.4. Errors and help. 21
5.5. End of Problem 23
5.6. Grading. 25

Appendices

A. The Exams Used in the Experiment 26

Sample Problem Displays. 29
Quirks of the Problem Generator/Graders. 34

B. Questionnaires 41

Background Sheet 42
Questionnaire on PLATO Exam and on Paper Exam. 43
Written Comments 47

C. Experiment Administration. 49

Activities During Experiment Session 50
Users Guide for PLATO Exams. 51
Time Sheet 52
Log Sheet. 52

D. Definitions of Categories of Activities. 53

E. Activities Log for Each Subject. 55

1. Introduction

This report provides details of an experiment conducted Monday, September 22, 1975. An analysis of the data is contained in a companion paper:

Hansen, W. J., R. Doring, and L. R. Whitlock, "A Videotape Analysis of Student Performance on an Interactive Examination," UIUCDCS-R-76-836, Department of Computer Science, University of Illinois, Urbana, Illinois (1976)

So it can stand by itself, this report repeats from the other sections 2 and 3, and the major figures in section 4.

2. Environment

The examination system in question - the Generative Exam System [Whitlock, 1976] - offers many advantages. A wide variety of question schemes are available. Geographic and temporal scheduling problems can be simplified by the network nature of the host system. The system can offer non-passive problems wherein subsequent subproblems reflect prior performance. Each student receives a slightly different form of each question (to encourage honesty). Finally, scores and correct answers are instantly ready for student review, and scores and statistics are instantly ready for the instructor.

The exam system, in turn, is but one component of the Automated Computer Science Education System (ACSES) under development by the Computer Science Department at the University of Illinois [Nievergelt, 1976]. ACSES includes over 100 instructional lessons, some twenty of which are in regular use for over 3000 introductory level students each year. These students constitute the population for which the exam system is designed. They are non-major underclassmen from a wide variety of fields. They have undeveloped typing skills, minimal exposure to interactive systems, and little motivation to learn computer programming.

ACSES is implemented on the PLATO system for computer aided instruction developed by the Computer-based Education Research Laboratory, also at the University of Illinois [Bitzer, 1973]. The University's PLATO system is currently the most fully used; it has 1000 terminals connected and about 400 are usually in operation. From a terminal a student can have access to several thousand hours of instruction, several hundred of which have been polished for - and used in - regular courses.

Though the PLATO system has many advantages, it imposes serious constraints on processing power, memory size, disk accesses, and display speed. Each user is limited to a maximum of ten "TIPS" (thousands of instructions per seconds) and good response cannot be expected if a user requires more than three TIPS. This is adequate for simple question-and-answer interaction, but it cannot support sophisticated versions of data base search, program text analysis, or exam grading. Program and data memories are limited, so large programs and data bases are impractical. Core limitations could be avoided by retrieval from disk, but each user is restricted to an average of one disk access per minute. In the Generative Exam System, this resource is consumed in reading and updating student records for each switch from one problem to another.

PLATO employs a plasma panel terminal. Its 512x512 dot display is high precision but, like a storage tube, cannot support rapid animation. Because it is driven by 1200 baud lines with a maximum of 180 characters per second, the terminal takes ten seconds to display a full page of text. Text may be written in a very rich character set because the terminal provides 128 built-in characters and memory space for another 128 to be defined by the user program. Unfortunately, it can take over ten seconds to transmit the codes to define any substantial number of extra characters.

The keyboard is an augmented typewriter design with a number of "function" keys, including NEXT, BACK, DATA, and HELP. Though a lesson designer can specify any response to these keys, they have certain conventional uses. HELP usually causes display of some additional explanation. BACK moves the lesson back to the last section of material covered. Shifted-DATA often returns the display to the index page for the lesson.

NEXT has two possible meanings: sometimes it terminates an answer and sometimes it simply signals that the student is ready to go on. Usually when the student answers a question and presses NEXT the system responds with "no" or "ok" following the response. This feedback is an important reinforcement during instruction, but must be switched off for exams.

The Generative Exam System is structured as a central monitor and a collection of "problem generator/graders" (PG/G's). The central monitor displays the exam cover page, transfers to individual PG/G's, stores student answers and scores, and generates statistics. Each PG/G displays one or more problems and interacts with the student to accept the answers. The generality of this structure supports an unlimited variety of question types and within each PG/G random generation techniques provide an infinite number of questions.

Each subject took two exams, one form A and one form B. Each form used the same four PG/G's:

1) Arithmetic expressions - 12 variables and their values were displayed and the student was asked to evaluate five expressions involving those variables.

2) FORTRAN syntax - This PG/G included a page of instructions, a cover page, and three problem pages. Each problem page displayed a FORTRAN statement with possibly a syntactic error created with an Extra, Missing, or Replaced basic symbol. For the experiment, form A had assignment, ASSIGN, and PRINT statements, while form B had assignment, GOTO, and DO. In each case the answer was to be given by specifying a type of error (E, M, R, or None) and the associated basic symbols.

3) Print with FORMAT - A print statement and a FORMAT were displayed along with a grid for the answers. The student entered a line of output and then specified where in the grid it should go. Form A had three F format items; form B had three I format items.

4) DO loop - A DO-loop with a PRINT statement was presented and the student had to specify the values that would be printed. This PG/G

graded interactively; each answer was checked when it was entered. Points were immediately deducted for an incorrect answer and a second chance was given. This scheme tried to avoid the problem of propagation of errors.

3. Experiment Design

A two by two design was chosen with subjects in each cell taking two forms of an exam - one on paper and one on PLATO. Both forms were generated by the system and copied on a screen copier to generate the paper version. The design assigned exam forms and methods as shown in figure 3.1. Because the two forms were considered essentially equivalent (and turned out to be so), the subjects were split between PLATO-first and paper-first so neither treatment was favored by the subjects' self-expressed typing ability. Having only one set of videotape equipment and one set of proctors the experiment was run in separate sessions of roughly one hour each.

Subjects were recruited in an elementary computer science course similar to those for which the exam system will be used (Computer Science 101, an introduction to FORTRAN for engineering students, Fall, 1975). They were offered the inducement of practice for the hour exam they would be taking a week later. For fairness, however, the experiment exams were made available to all students after completion of the experiment. (More recently, the exam system has become a popular way to study for exams.) We decided to tape only four subjects for several reasons: the major anticipated effect was large (i.e., slowness on PLATO), the equipment was expensive, and we needed experience. We had enough volunteers to schedule an additional four subjects who were not videotaped; this proved fortunate, because one of the eight subjects

failed to appear at the appointed hour. When students volunteered, they were asked to state their typing ability.

Taped Subject	Untaped Subject	Method and Form	
		First	Second
SA		PL A	pa B
SB	SE	pa A	PL B
SC	SF	PL B	pa A
SD	SG	pa B	PL A

Figure 3.1. Experiment Design. For this paper, the subjects have been randomly coded as SA,..., SG. "PL" = PLATO. "pa" = paper.

Each experiment session began with a practice exam on PLATO, followed by the two exams dictated by the design. The post-test questionnaire elicited reactions to the two methods of giving exams.

The video camera was located slightly behind the subject's shoulder so as to be out of sight while still recording the subject's face and hands, and the general appearance of the PLATO screen or paper exam. Due to low resolution, it was not possible to record the details of the work, so an observer was positioned behind the subject to make a manual record including the sequence in which the problems were worked. A clock was positioned beside the terminal so the time was recorded. To avoid time pressure, the clock was turned so the subject could not see its face.

Data analysis began by coding each "activity" observed into one of the categories listed in figure 3.2.

Brief Title	Code	Description
Think	RT	Read and Think
"	CP	Calculate with Paper and Pencil
Answer	EA	Enter Answers
Select	PS	Problem Selection
Generate	PG	Problem Generation (PLATO only)
Load	LC	Load Character Set (PLATO only)
Display	PP	Problem Presentation (PLATO only)
Trouble	WN	What Next (subject confused)

Figure 3.2. Activity Categories. For most of this paper, CP--calculate with Paper and Pencil--has been lumped together with "Think."

4. Results

Data for all seven subjects are presented in Figures 4.1 and 4.2. The former shows both the subjects' backgrounds and their performance on the experiment. One conclusion from this data was that the differences between taped subjects were exacerbated by the experimental situation. This is echoed on the post-test questionnaire results shown in 4.2.

Details of the taped subjects' behavior appear in Figures 4.5 and 4.7 (Numbered for the analysis paper. Omitted figures present details of the analysis.) The first section of 4.5 presents total times in each category for each subject. The second section lists the number of activities corresponding to the times above. The third section summarizes the data in several categories. A major conclusion from this data was that the PLATO overhead on an exam need not be more than twenty percent of the time to take the exam on paper.

Figure 4.7 shows the time each subject spent in productive work on each problem. One observation is that subjects reviewed much more on PLATO, possibly because of lack of confidence in system reliability.

(a) Background Data

Subject	Year	Field	CS 101 Self-Evaluation	CS 101 Expected Grade	Typing Skill	Approx. Hrs. On PLATO	GPA	First Hour Exam
SA	SO	Cer.Eng.	Good	B	Medium	20	B+	A
SB	JR	Elec.Eng.	Good	B	Slow/sure	3	A-	B
SC	SR	Chem.	Good	B	> Good	3	B-	B-
SD	SR	Met.Eng.	Fair-	B	Slow/sure	10	B	B
SE	SO	(LAS)	Fair	C	Slow/sure	10	B	C
SF	JR	Math.	Good	B	Slow/sure	5	B+	B
SG	JR	Physics	Very Good	A	Slow/sure	>>20	A-	A

(b) Results

Subject	First Exam	Total Time (Min) Prac.	Total Time (Min) PLATO	Total Time (Min) Paper	Total Score PLATO	Total Score Paper	Pts/Min PLATO	Pts/Min Paper	Satisfaction
SA	PL A	8	23.2	7.3	25	30	1.1	4.3	1.7
SB	pa A	13	19.8	14.8	24	35	1.2	2.3	3.2
SC	PL B	8	26.7	10.8	27	25	1.0	2.3	3.6
SD	pa B	5	17.9	11.4	38	39	2.1	3.5	3.9
SE	pa A	9	24	14	26	26	1.1	1.9	3.2
SF	PL B	8	16	10	47	36	2.9	3.6	2.4
SG	pa B	11	15	21	49	50	3.3	2.4	3.8
average		8.9	20.4	12.7	33.7	34.4	1.8	2.9	3.1
standard deviation			4.1	4.2	10.1	8.0			.7

Figure 4.1. Summary of all Subjects. For confidentiality, "GPA" and "score on first hour exam" have been converted to letters. Under "first exam", PL = PLATO, pa = paper, A = form, and B = form B. "Satisfaction" is a weighted average of post test questionnaire responses (5 = complete satisfaction with PLATO exam).

	SA	SB	SC	SD	SE	SF	SG
1. liked PLATO vs. paper (e = much less)	d	c	d	a	c	d	c
2. experiment bother (c = no bother)	a	c	b	b	c	b	c
3. PLATO content difficulty (e = too trivial)	c	c	c	c	c	d	d
4. PLATO instructions difficulty (e = very confusing)	e	c	c	d	c	d	b
5. concentrate on PLATO (c = not able to)	b	a	a	b	b	a	a
6. paper content difficulty (e = too trivial)	c	c	c	c	c	d	d
7. concentrate on paper (c = not able to)	a	a	a	a	a	a	a
8. PLATO switch frequency (c = once each)	a	b	b	b	c	b	c
9. PLATO switch difficulty (c = very little)	a	c	c	c	b	b	c
10. paper switch frequency (c = once each)	b	b	b	c	b	c	c
11. PLATO keyboard hindrance (e = keyboard much preferred)	b	b	c	e	c	c	a*
12. PLATO typing impact (d = no influence)	b	c	c	c	c	b	b**
13. PLATO reading time (e = much less)	b	c	c	b	c	c	c
14. PLATO vs. paper preference (a = PLATO; b = paper; c = both; d = don't care)	b	d	c	a	a	b	d
15. did PLATO delay (yes; no)	y	y	n	n	y	y	y
Satisfaction	1.7	3.2	3.6	3.9	3.2	2.4	3.8

Figure 4.2. Summary of Post Test Questionnaire Responses. The taglines in parentheses indicate the permitted range of response and the meaning of that response. *Response "a" meant "no hindrance". "b" through "e" expressed preference. **The subject wrote that the impact was "increased speed and reduced clerical errors".

times

sub-ject	first	Think PL	Think pa	Answer PL	Answer pa	PLATO Activities Generate	PLATO Activities Display	PLATO Activities Load	Select PL	Select pa	Trouble PL	Trouble pa	Total PL	Total pa
SA	PL A	12:07	6:22	1:13	:34	:33	1:17	:20*	2:13	:23	5:28	--	23:11	7:19
SB	pa A	12:27	13:02	1:43	1:29	:31	:43	0*	1:42	:08	2:40	:08	19:46	14:47
SC	PL B	18:43	9:41	:36	:49	:26	1:33	:19*	2:34	:17	2:33	--	26:44	10:47
SD	pa B	10:19	8:55	1:27	1:19	:28	1:34	:18*	2:42	:23	1:04	:44	17:52	11:21
average		13:24	9:30	1:15	1:03	:30	1:17	:19	2:18	:18	2:56	:13	21:53	11:03

activities

SA	PL A	34	26	24	17	3	14	1*	17	9	13	--	106	52
SB	pa A	31	27	22	23	3	8	0*	12	7	8	1	84	58
SC	PL B	39	36	21	17	4	16	1*	20	12	5	--	106	65
SD	pa B	40	22	28	19	4	16	1*	22	7	3	3	114	51

summaries

		pts PL	pts pa	Efficiency (pts/prod sec) PL	Efficiency (pts/prod sec) pa	Productive (Think+Answer) PL	Productive (Think+Answer) pa	PL Overhead (Display+Gen) PL	Reducible (Load+Select) PL	Reducible (Load+Select) pa	#S Activities (Thk,Ans,Sel,Trbl) PL	#S Activities (Thk,Ans,Sel,Trbl) pa	# Prob PL	pgs pa
SA	PL A	25	30	.031	.072	13:20	6:56	1:50	2:33	:23	88	52	11	9
SB	pa A	24	35	.028	.040	14:10	14:31	1:14	1:42	:08	73	58	8	6
SC	PL B	27	25	.023	.040	19:19	10:30	1:59	2:53	:17	85	65	13	11
SD	pa B	38	39	.054	.064	11:46	10:14	2:02	3:00	:23	93	51	13	6

Figure 4.5. Summary of Detailed Times for Taped Subjects. Times are in minutes and seconds. *The character set for P2 was preloaded during the practice exam so the times and counts shown are only for review.

	SA (PL A)			SB (pa A)			SC (PL B)			SD (pa B)				
	first try	review	pts	first try	review	pts	first try	review	pts	first try	review	pts	max pts	avg pts min
1. Exprs	2:29	1:03	15	5:06	--	9	4:31	1:27	12	2:25	1:48	9	15	2.4
2. Syntax 1	:22	:07		:13	:06		:43	:39		:16	:22			
" 2	1:30	1:18	10	:48	--		1:15	:23	15	:36	:05	10	15	3.7
" 3	:07	:07		1:35	--		1:33	:29		:15	:47			
3. PRINT	:27	--	0	1:26	:40	0	1:35	1:08	0	2:27	:35	9	10	1.1
4. DO	5:45	:05	0*	4:16	--	0*	5:31	:05	0*	1:59	:11	10	10	.6
total	10:40	2:40	25	13:24	:46	24	15:08	4:11	27	7:58	3:48	38	50	

paper

1. Exprs	1:57	:16	12	4:01	--	9	3:31	:55	12	3:18	--	6	15	2.8
2. Syntax 1	:20	--		1:28	--		:46	:07		:27	--			
" 2	:19	:14	10	3:08	--	11	1:54	:13	6	:27	--	15	15	2.9
" 3	1:23	:35		:38	--		:56	:09		1:18	--			
3. PRINT	:11	--	0	3:32	--	5	:06	:02	0	1:26	--	10	10	2.8
4. DO	1:41	--	8	1:44	--	10	1:51	--	7	3:18	--	8	10	3.9
total	5:51	1:05	30	14:31	--	35	9:04	1:26	25	10:14	--	39	50	

Figure 4.7. Productive Work Times for each Problem. Times shown are the sum of Think and Answer times in minutes and seconds. Subjects seldom reviewed a question more than once. *Log shows one or more correct answers but grading algorithm subtracted all points.

5. Forty Design Proverbs

Our experienceswith the exam system and the videotape experiment have led to development of design principles that minimize user difficulty. These principles show some similarity to those in

Hansen, Wilfred J., "User Engineering Principles for Interactive Systems," Proceedings Fall Joint Computer Conference V. 39, AFIPS Press, Mondale, NJ, pp. 523-532.

Many of the "proverbs" in subsequent sections are particular instances of a few *meta-proverbs*.

01. Be Predictable. Users experience enormous frustration when a system does not behave as they have come to expect it to behave. This proverb is especially hard to follow, and, therefore, especially important in a multi-component system like the exam system with its many independently written PG/G's. Areas in which the system should behave predictably are similar to those covered in the next proverb.

02. Follow the Standards. Observance of predefined standards reduces design effort and gives an air of familiarity to the product which reduces the users effort to learn how to control it. For the exam system standards exist for several areas: control key conventions, problem page appearance, the response to an answer entry, and the order of generating displays. If we knew enough, we would even specify standards for problem wording.

03. Provide Reversibility. For each action the user can take there must be an obvious way to undo that action. This provision reduces the pressure on the user by reducing the penalty for typing and procedural errors. For the exam system typical actions are to enter an answer or move to the next problem. The reverse actions are to delete an answer and to move to the previous problem; both must be available at every point in every PG/G.

04. Keep Instructions Simple. If the procedures cannot be explained simply, they are too complex and must be revised. A surfeit of instructions reduces the emphasis on each to the point where even important ones are forgotten.

Many of the principles relate to the standard control keys so it is necessary to discuss them here. There are five control keys--NEXT, BACK, HELP, DATA, and LAB--and alternates with the shift key--shift-NEXT, shift-BACK, shift-HELP, shift-DATA, shift-LAB.

NEXT--Normally has three uses--end of an answer, end of reading a paragraph (add some to page), and end of reading a page (go to next page). The user usually cannot predict the distinction; often the end of an answer will cause a new page.

Exam system--End of answer or move the arrow to the next answer on this page. The arrow cycles from the bottom answer back to the top one.

shift-NEXT--No usual function. Some Computer Science lessons skip to next topic. Some non-Computer Science lessons jump to the table of contents.

Exam system--Go to the next problem page.

BACK--Usually takes the user back to the previously displayed page, perhaps for review.

Exam system--Move the arrow to the previous answer slot. Cycle top to bottom.

shift-BACK--Usually goes to beginning of previous "major" section of the lesson (or back to table of contents from first section).

Exam system--Go to previous problem page.

HELP--Normal lessons provide a response to the HELP key whenever the user may have difficulty understanding the material or answering a question. Sometimes HELP invokes a new display and BACK is used for return.

Exam system--HELP should explain restrictions on the answers. It should never invoke a new display.

shift-HELP--Often used to provide help on key conventions. At least that is its use in the exam system.

DATA, LAB, shift-DATA, shift-LAB--Used for many purposes, few of them mnemonically suggested by the names of the keys.

Exam system (and other Computer Science lessons)--shift-DATA transfers the user to the exam cover page. The others are used for various purposes by individual PG/G's, for example, LAB provides access to a calculator in one and LAB/DATA move a pointer in another.

5.1. Instructions and Preparation

Even a crystal clear design is futile if students are not aware of its facets. These proverbs suggest ways to aid the student in understanding the system.

11. Provide practice exams. These will help students learn the conventions and gain confidence that the system retains answers and computes a grade only on the final answer. Quizes at the ends of lessons should follow similar conventions so the student is at least accustomed to the fact that the system moves each answer after it is entered. (Unfortunately, many of our quizes predate this finding and use other conventions.)

12. Provide written instructions. Prior to the practice and actual exams the student should receive a printed list of instructions for preparation and for reference while working problems. The instructions should emphasize the freedom of, and mechanisms for, moving around in the exam and changing answers.

13. Stress the mechanics of moving from problem to problem. If students are aware of how to move between problems, they can always escape from trouble by going to another problem. The instructions for our experiment (Appendix C) stressed the possibility of moving, but failed to say how.

14. Stress the mechanics of entering and changing answers. This is less important than proverb 13 because students will enter more answers and will get more practice. Nonetheless, some stress is needed because the conventions differ from instructional lessons.

15. Don't patronize. Assume the student is intelligent even if not informed about the details of some concept of procedure. Do not present disparaging remarks in case of error. Avoid cute remarks. The subsystem through which students enter the exam system greeted the finished student with a gratuitous "Congratulations for completing csxmonitor!" This is especially inappropriate if the student "completed" the exam by leaving some problems undone and scoring poorly.

16. The proctor must take the exam and make mistakes. Only with experience will the proctor be able to help students when they get in trouble. At worst the experience should help the procotr know how to get around a previously undiscovered bug in a PG/G.

17. The proctor should avoid random roaming of the exam room. Students cannot profitably use the answers from other students, and roaming may arouse the sensitivity of those who prefer not to display their ignorance.

5.2. Problem Generation and Display

21. Use standard page layout (see Figure 5.2.1.). This will provide a common, professional appearance, disguising the polyglot authorship of the PG/G's and providing continuity of expectation between the problems. All learning that carries over from one PG/G to the next saves the student intellectual effort.

22. Use a standard display order (see Figure 5.2.1.). As with 21, this increases commonality. In addition, it saves time because the student can read important information while more is being displayed, and because generation time will overlap with display time. (Central processor execution continues as soon as display commands have been transferred to a buffer for subsequent transmission to the terminal.)

23. Put multiple problems on a page. In this way a student can do more work with each overhead penalty. The display time for the problems might be the same, but there is display overhead in headers and key instructions and there is "context switch" overhead in recalling the meaning of a problem format. (However, care must be taken to avoid a cluttered appearance on the page.)

24. Specify clearly the set of possible answers. If the PG/G may reject answers and if the form of the answer is not obvious, some statement of the expected form is needed. Preferably though, the question will be revised to the point where the expected form is obvious.

25. Use standard problem wording. While it is clear that the wording of a problem can seriously influence a student's ability to answer it (especially for foreign students), it is entirely unclear how to best word questions. This must be left for other research.

26. Do not use special character sets. The overhead to load such character sets is unacceptable in an exam. It is acceptable to load as many as five special characters for special applications. Unfortunately, there is not even a good way to disguise the load time because characters should be loaded before beginning the display in order to avoid a disconcerting pause during the display process.

27. Control difficulty differences. It is essential that two students taking the same exam at the same difficulty level receive problems of nearly equal difficulty. To do otherwise is to raise serious questions of fairness. In order to control difficulty, the designer must carefully list all the problem components which can affect understanding; these include expression complexity, statement order, and even such trivia as the length of identifiers.

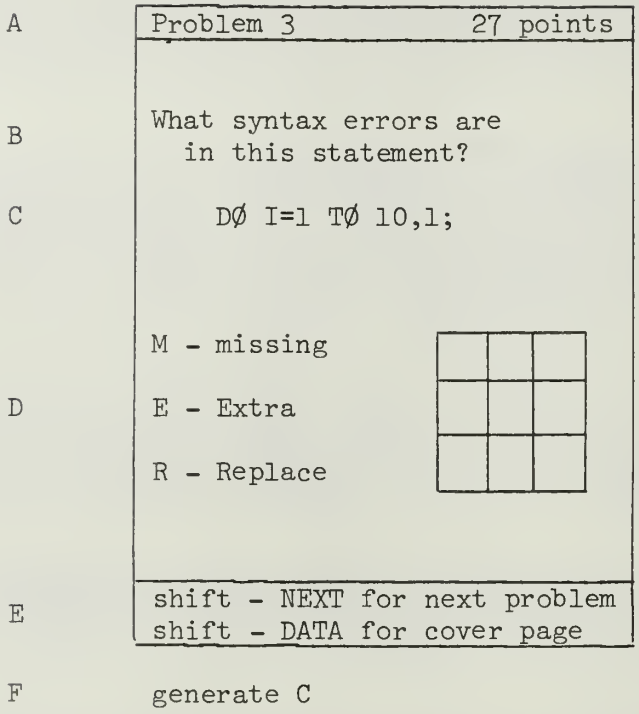


Figure 5.2.1. Standard Page Layout and Display Order. If problem generation is fast, the display should be in the order B, F, C, D, A, E so the generation will occur while B is being displayed. For slower problem generation the order should be B, A, E, D, F-C where F-C implies piece-by-piece display. If there is initialization for F, it should start as early as possible.

5.3. Entering Answers

31. Follow the standard control key conventions. In this context, the most important conventions are those which specify that NEXT and BACK move the answer arrow up and down. These must be followed so the system behavior will always be predictable when the student is at the most vulnerable--i.e., entering answers and trying to remember the answer long enough to get it completely entered.

32. For every student action provide a system reaction. Though it may be sufficient to respond to an answer entry by simply moving the arrow to the next position, it is probably better to also move the answer. This indicates clearly that the system has accepted the answer. Other actions that deserve response are all key presses, however, it is PLATO convention to simply ignore illegal presses of the control keys so this is a reasonable procedure to follow.

33. Use the standard scheme for changing answers. This follows from the previous proverb. The answer has already been moved to the answer slot and is readable while a new answer is being entered. It is important to note that if simply NEXT or BACK is pressed, the existing answer is unchanged. To erase an answer the student must type a blank and then press NEXT (or some other control key).

34. Use the system "arrow" command for input. This command provides a number of features like erasure and editing which the more sophisticated students use advantageously. It is also faster than any other scheme of entering answers (since the only other scheme on PLATO is to accept characters one at a time and invoke the program for each character). A corollary is that the standard system "arrow" character should be displayed before the answer slot. (One of the experiment PG/G's used another arrow shape and this caused some very small confusion.)

35. Accept answers only when a control key is pressed. This proverb forbids the practice of accepting a short or one letter answer as soon as it is typed. Doing so causes two difficulties. First, the habit of not pressing NEXT is reinforced. Subjects sometimes forgot to press NEXT when it was required and thus wasted time. Second, and more critical, the student loses a chance to review the answer before making it final. Even though the answer can be changed, the student is likely to perceive a loss of control over the behavior of the system.

36. Always accept answers when a control key is pressed. On several occasions during the experiment and our personal use of the system, the correct answer was entered and then some control key other than NEXT was pressed. This resulted in an ignored answer and a loss of points. This problem can be avoided without great difficulty using facilities provided by the TUTOR language.

37. Do not allow the user to use special characters. Users should not be required to use special characters (using FONT or MICRO) in answers because special learning is needed to know how to enter those characters. The proverb is stated in this stronger version so that authors do not even consider the possibility of special characters (which are discouraged by proverb 26 anyway).

38. Accept answer lists. In the DO loop problem and others, more than one answer is to be placed on a line. Sometimes a student will enter both answers at the arrow even though the arrow will move to the adjacent slot for the next answer and ignore the answer already supplied. To provide for such eager students, PG/G's must take the special effort required to capture all the information entered at every answer slot. For example, if more than one number is entered and there is more than one slot on the line, the answers should be spread across the slots. This proverb also requires the production of error messages when the student enters answers that will not fit into slots on that line. Extra characters should not be ignored as they can be with standard PLATO mechanisms for accepting numeric answers.

5.4. Errors and Help

41. Avoid rejecting responses. Many problems with response rejection appear in the subject logs (Appendix E). Responses are not rejected on paper and students are disturbed when a response is rejected.

42. Give prominent feedback on response rejection. If response rejection is unavoidable, the feedback must be large enough to see and should be placed on the screen close to where the student can be assumed to be looking. A message can be highlighted by flashing it as long as the flashing routine is not deaf to keys (i.e., any keypress should terminate flashing and proceed).

43. Give precise feedback on the reason for rejection. A simple "no" is not enough. The message should state what is wrong and even suggest how to fix it: "Do not use 'E', move the decimal point," "Specify a replacement character from (, -, *, ,, ., =".

44. When a response is rejected, do not go to a special state. Many subjects had trouble with the "Replace" option in the syntax problem; another lost points because of the irreversible nature of the DO loop problem. After rejection the system should be in exactly the same situation as before the attempt, except that the rejected response should be visible and proverb 45 should be followed. The terminal should behave as a problem-independent, answer-slot-filling device.

45. Arm the HELP key when a response is rejected. If the student presses HELP immediately after response rejection, the assumption is strong that information concerning the rejection should be offered. Consecutive presses of HELP should elicit more detailed information, although this must be limited by the time available to implement the problem generator.

46. Use on-page HELP when possible. For both ordinary help and response rejection help, it is preferable to display the help information on the same page as the problem even if this requires erasure and regeneration of up to a quarter of the screen. By supplying help on the same page, the student will be able to refer to the parts of the problem while reading about them and will not suffer the context-switch delay when returning from a help page to the main problem.

47. Avoid designs that rely on the HELP key to assist the student. Such designs are certain to be too difficult; they need more work to reduce their complexity.

48. Since the HELP key is not to be used, use it copiously. Use HELP to explain the obvious even items that are already explained on the page. Take care, however, against confusing extra words with extra explanation. Explanation is only valuable if the student can find it.

49. Do not assume the student has read and remembered the instructions. There are so many things to remember (the facts being tested, the particular instructions for the question, and the general procedure for using PLATO for the exam) that the student cannot initially remember them all. As a consequence, the system should not be harsh with students who have made some mistakes.

4A. The proctor must give explicit aid. On early experimental sessions we attempted to help students only minimally; i.e., by pointing to the instruction that would solve their problem. This is unsatisfactory. When a student turns to a human for help, it is evidence of complete failure to find comfort in the system. The more quickly the human can explicitly help the student back on the right track, the higher are the student's prospects for comfortable interaction with the system.

5.5. End of Problem

On paper when a student finishes work on a problem, the work can be continued on the next problem directly. On PLATO, a few more considerations are needed.

51. Emphasize the last answer. Some of our PG/G's, for example, the arithmetic expression one, simply move the answer arrow back to the first answer slot after the last answer is entered. On one or two occasions we observed students exhibit befuddlement over the implied suggestion that they try all answers again. As an indication of the end of the problem, we suggest that before returning the arrow to the top the PG/G should cause the key convention message to flash. This message includes the information of how to move on to the next problem.

52. Avoid cover pages. Because of the time to display a page, PG/G's should avoid nonproductive displays such as cover pages. Most problems can be displayed entirely on one page, but some--like short-answer--are more convenient with one page per subproblem. The difficulty is that there must be some way to go to individual subproblems. Certainly access to subproblems must be possible via the simple shift-NEXT and shift-BACK used for transfer between problem generators. That may be enough if the number of subproblems is kept small. As an alternative, we have experimented with a vector of problem numbers displayed at the bottom of the page. The difficulty with this scheme was that response to the answer arrow could be either an answer (distinguished as alphabetic) or a number specifying a new subproblem. This added confusion to both answer entry and subproblem selection. It might be satisfactory, however, if the index vector were only displayed on command. An as yet untried alternative is to have a "mini"-cover page with just the problem numbers and the student's answers (the latter as a clue to the problem contents). In this scheme the small amount of text could be quickly displayed.

53. Adopt a standard end-of-answers convention. On paper exams it is usually assumed that a student has written all the answers felt to be needed. For example, on a DO loop problem the answers given will be graded. Occasionally a syntax error problem will provide the option of stating that there are no errors. Such exceptions are more difficult on PLATO because the student is more uncertain of how to enter answers; the system constrains answers in so many ways. Of the experimental problem generators the DO loops and syntax error problem both expected "non-answer" answers: "end" for the end and "N" for no errors, respectively. Ideally, problems with empty-list answers would be eliminated, but this is not always possible.

This proverb demands that there be a single "non-answer" convention for the system. It should be an "end-of-answers" code rather than an "empty-list" code because students can use the former even at the end of non-empty lists. However, even if the code is omitted from a non-empty list, the grader

must make the standard assumption that the student has entered all the answers the student felt necessary. A reasonable end-of-list code is "end". One refinement might be valuable: When a student attempts to leave a problem without making any entry in a list, the system might flash a reminder of the empty-list convention. Unfortunately, this conflicts with predicatable behavior because the system convention is to transfer to the requested problem immediately when a transfer key is pressed. One partial solution would be to flash the end-of-list message (which should already be visible in the problem statement) and then transfer to the requested problem after a very brief pause (say 2-3 seconds).

5.6. Grading

In one sense automated grading is absolutely fair since the same algorithm is applied to every student. However, the system could be inequitable to the extent of basing the grade either more or less on the ability of the student. The proverbs in this section are intended to suggest how grading can more accurately reflect student ability.

61. Give partial credit. The "precision" of a grading algorithm can be measured by the number of possible grades it can assign. An all-or-nothing grade less accurately reflects gradations of ability than a grade that may take any of several intermediate values. Even some of the graders used in the experiment gave partial credit: For correct value with wrong type or sign on the arithmetic expressions, for incorrect print position, for correct second guess on the DO loop problem; but more can be done--such as proximity of value (especially in terms of character representation), and whether the judgment of presence of syntax error was correct. In most cases the amount of partial credit can and should be left to the judgment of the instructor when selecting options.

62. Use relative grading. In many problems an answer to some subpart depends on the previous answer. In manual grading of such problems it has been traditional to give credit for subsequent answers based on the previous value; thus, an incorrect value would itself be marked wrong but its effects would not propagate. This tradition can and must be continued with computer administered exams. Moreover, because of the problems of typing accuracy, subsequent answers that are correct in an absolute sense must also be accepted. This maxim applies in the obvious sense to noninteractive DO loop problems, but also applies to the print FORMAT problems. The latter grader judges the position of an answer as relative to the position of the previous answer. It must also perform a pattern match to determine which strings in the answer match which of the problem data.

63. Do not use interactive grading. Our subjects had fairly strong reactions to rejection of answers by the interactively graded problem on DO loops. One student lost points simply by failing to note the error message and supplying the second iteration when the program called for a retry of the first answer. Many proverbs, including "Provide Reversibility" are violated by interactive grading. Moreover, the benefit expected from interactive grading--reduced error propagation--can equally well be achieved by relative grading. If interactive grading is to be used, it should be isolated and carefully explained to the student. It is, however, possible that an entire exam of interactively graded questions would be acceptable, but this would presuppose construction of numbers of interactive questions, and the time is better spent devising relative grading algorithms.

Appendix A. The Exams Used in the Experiment

Two instances of the exam were used: form A and form B. When administered on paper, the questions shown here (p. 29-33) were presented; when administered on PLATO, slightly different versions were presented to each student. The two exams each had four problems:

1. Evaluate five FORTRAN expressions (the system chose five from a collection of about 20).
2. Find syntax errors in three statements; each statement was presented in a separate display. (Form A had assignment, ASSIGN, and PRINT. Form B had GOTO, assignment, and DO.)
3. Show the output of a PRINT and FORMAT statement. (Form A used three F-format items, Form B had three I-format items.)
4. Show the output of a PRINT in a DO loop. (This problem generator/grader attempts to provide interactive grading so an error on one value does not mean loss of points on subsequent values. In addition, the generator creates many different variants, possibly involving subscript-out-of-bounds and modification of values used by later steps.)

The practice exam consisted of two expressions from 1 above and one syntax error from 2. The following pages illustrate the question formats used in the exam; they are taken from the paper version of Form A. The actual paper exams differed as follows:

Form A: As shown except problem 2 (FORTRAN syntax) had two more pages with these FORTRAN statements:

12.....ASSIGN.25.TO.WUZXIYHJ

1......PRINT.,X,XQUX

(The dots indicate spaces; they are one-fourth the size of a period.) The ASSIGN was a great trial to the students because (a) they had not previously encountered an ASSIGN statement, and (b) "extra HJ" was not an answer that would have been accepted by the problem generator/grader. (Generation of the 8-letter identifier was a bug.)

Form B: Problem 1. Using the same variable values, the expressions were

N+IFIX(S)/I, B**L**3.0, (S-B)*C/4,

A/C/B*3, -B**C

Problem 2. The three statements were

68.....GO..TO.778

3......AL=MKFT

41.....DO.7.JFM=9*31

Problem 3.

```
    I=68
    J=8
    K=2
    PRINT 20,I,J,K
20  FORMAT('1',2I4,I5,I5)
```

Problem 4.

```
      INTEGER M(6)
      READ,M
      DO  20  I=1,8,2
      IF(I.EQ.6)GO TO 20
      M(I)=M(I)-M(3)
10    PRINT,I,M(I)
20    CONTINUE
```

Values for M: 4, 3, 6, 4, 4, 6

This is harder than form A because M(3) is modified before the last iteration.

COVER PAGE

Exam number 15 for course 'csa'; for practice.

Number	Keyword	Weight	Score
1	Fortran expressions	15	
2	Fortran syntax	15	
3	Print with Format	1Ø	
4	Fortran DO loops	1Ø	
	TOTAL	5Ø	

Problem 1 Fortran expressions 15 points

Type in the value for each expression. Include a decimal point if and only if the value is real:

A=18.	B=3.Ø	C=2.	S=-4.5	Y=1Ø.Ø	Z=5.
I=2	J=3	K=4	L=1	M=12	N=8

A-C*Z-B

ABS(-B-A/S)

-B**I

B**L**3.Ø

(S-B)*C/4.

FORTRAN Syntax errors Question # 1 of 3

Find all the syntactic errors in the following FORTRAN statement

```
5.....YN=CD-33/D
```

[E] EXTRA

[M] MISSING

[R] REPLACE

[N] No errors

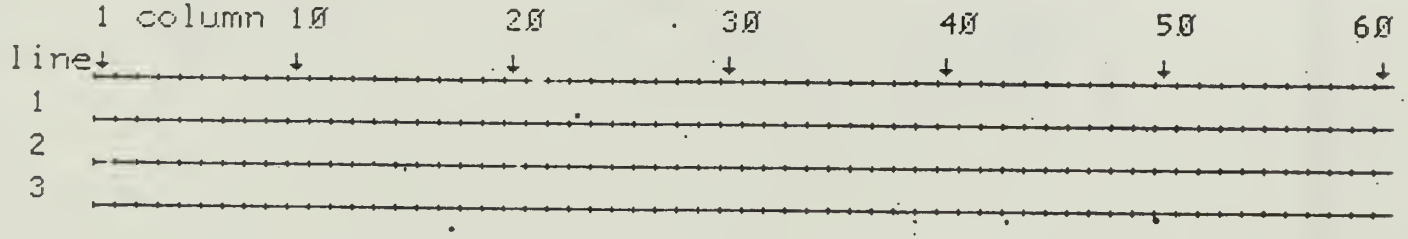
→

errors		symbols involved	
#	type		
1			
2			
3			
4			

Problem 3 PRINT with format 10 points

Show exactly what is printed by this program segment.

```
      A=7.3862
      B=7.85
      C=8.74
      PRINT20,A,B,C
   20 FORMAT('1',F5.2,2X,F4.1,F3.0)
```



Problem 4 Fortran DO loops (10 points)

What does this Fortran fragment print?

```
      INTEGER M(6)
      READ,M
      DO 20 I=1,6,2
      IF (I.EQ.3) GOTO 20
      M(I) =M(3) + 1
10    PRINT, I, M(I)
20    CONTINUE
```

Output:

I M(I)

>

Initial contents of array:

	subscript					
	1	2	3	4	5	6
value	2	0	1	1	1	2

There is art to creation of a successful Problem-Generator/Grader. The four used in this experiment illustrate many of the possible difficulties.

P1. FORTRAN Expressions (author: L. R. Whitlock)

This PG/G was designed from the start as a makeshift. It provides only a standard set of 12 variable definitions, and a small library of expressions. The only "generation" is to select expressions from the library at random. Since the experiment, a more sophisticated generator has been implemented.

Only two features of this PG/G annoyed subjects. First, it rejects answers using "E" for exponent, insisting, for example, that 1.5E2 be written 150. This is reasonable--especially since no variable value has an "E"--but it raises the question as to what is being tested. Second, the arrow moves from the last answer slot back to the first as discussed in proverb 51.

P2. FORTRAN Syntax (author: F. J. Izquierdo)

This problem-generator/grader was the result of the Master's project described in

Izquierdo, F. J., "A Generator/Grader of Problems about Syntax of Programming Languages to be Used in an Automated Exam System," UIUCDCS-R-75-755, Department of Computer Science, University of Illinois, Urbana, Illinois (1975).

A very important contribution of this effort was to show how to do a reasonable job of generating syntax errors without generating ambiguously correctable constructs. The approach is to generate errors by first generating a correct statement and then modifying symbols chosen from a certain set. By careful choice of this set the system avoids most possibilities for ambiguity.

The generator was completed before the exam system and independent of it, so any inconsistencies between the two are not mistakes but are simply due to different choices. This discussion will nevertheless point out the differences because they illustrate the kinds of differences which can cause annoyance to the user. Such differences contribute to nonpredictability and thus violate proverb 01.

a. The most serious difficulty with the syntax problem was a design error--when an answer was rejected a mode was entered wherein only a valid answer could be supplied. This occurred after entering "E" and "R" as the error type and then trying to specify some character not on the set of modifiable symbols. Almost every subject had trouble with this feature and had to be helped out verbally. (The problem was compounded by the fact that the messages describing key conventions were visible even though those keys were inactive.) See proverb 31.

b. A second area where subjects--at least all the videotaped ones--had difficulty was on trying to move from one subproblem to the next and trying to leave the syntax problem. Within the PG/G, shift-NEXT always went to the next subproblem, but after the last subproblem it returned to a syntax problem cover page. On the latter, shift-NEXT returned the student to the first syntax subproblem. This has been fixed by arranging the conventions such that the syntax problem appears to be consecutive pages with the last subproblem preceding the next problem page on the exam. See proverb 52. (E.g., the pages on the experiment would be exam-cover, P1, syntax-cover, Q1, Q2, Q3, P3, and P4; and shift-NEXT and shift-BACK move forward and back in this list.)

c. Unfortunately, the difficult goal of providing good problems was only partially met. Many examples of incorrect problems have been noted:

the eight letter identifier on form A,

generation of a statement longer than the space available,

generation of a not-yet-introduced construct (CALL) as the object of a logical IF,

a legal statement resulting from an intended error (DO 10 I=9*31)

implied semantic errors -- DO 25 KD = 93 13 (the bounds would have to be in ascending order: 9, 313)

statements with ambiguous corrections--DO 10 I = 6*, 1 or INTEGER I(5/ (initialization or dimension)

d. One way to solve the problems of (a) and (c) would be to revise the solution technique: the student would enter a new version of the line and it would be graded on its proximity to the original and its lack of syntax errors.

e. A less easy to solve problem is the generation of problems of very diverse difficulty. Some had no errors and some had only very trivial errors. Exhaustive study would be required to determine a hierarchy of difficulty suitable for doing better than this generator. Perhaps in the meantime this generator is suitable as a drill device rather than an exam device.

f. This was the only generator to delay subjects by loading a special character set. It did so because it uses the conventions (and some of the internal mechanisms) of the ACSES compiler system, which uses the same special character set.

g. In examining the logs, it can be noted that three activities were recorded for displaying a syntax problem. These were Present, Generate, and Present; the generator already follows the practice of displaying the basic problem statement before going ahead to generate the problem (though not before loading the character set).

h. Like the compiler system the syntax problem displays blanks in statements as a small dot (one plasma dot where a period is four dots). This led two of the videotaped subjects to ask whether the dots were the errors. Conversely, another subject inquired whether the error might be that the characters were in the wrong columns, so some statement that columns are correct is necessary.

i. There were a few trivial details in which the syntax problem differed from the rest of the system:

its cover page does not display the problem number of the syntax problem relative to the exam,

a problem yet attempted is marked with an asterisk on the syntax problem, while the same mark is used to indicate an attempted problem on the exam cover page,

the problem uses a different graphic for the answer entry arrow.

P3. PRINT FORMAT Problem (author: L. R. Whitlock)

Few students attempted this problem because the material had not been covered in lecture by the time of the experiment.

a. The generated FORMAT always includes a '1' as the first item, but there is no way to indicate that a line is to appear at the top of the page. Either the '1' should be deleted or one of the output lines should be marked as being the top of the page.

b. There is no control over how many test values will require rounding up before they are printed. In consequence, there are considerable differences in level of difficulty; for example, form A required rounding up three times while some PLATO versions of the problem did not involve rounding at all. (Indeed, at least two of the seven subjects lost points solely on failure to round up.)

c. This PG/G was one of the first to feature relative grading and partial credit. If the answer given does not exactly match the correct answer, the PG/G looks for the correct value in a wrong position and subtracts only a small number of points if a match occurs. If the answer still does not match, the fractional part is dropped and a match on the integer part is sought; again a success leads to only partial deduction. With this algorithm, a student will usually not lose full points for failure to round up (unless the rounding carries into the integer part as happened on one example).

d. The design of the answer entry scheme for a print FORMAT problem is difficult. Originally this PG/G displayed an extra line (with the tic marks above and below it) to enter a line for the output grid and a question as to which grid line the user wished to fill. After specifying a line number, the user entered text in the extra line and pressed NEXT; then the system transferred the contents of the extra line to the specified position in the grid. This scheme caused difficulty for at least SD and SG. It has since been revised so the user is asked what line to fill before the extra line is displayed. In this way, attention is first directed to the grid; then when a line is selected the means to fill it is provided. Yet another alternative would be to put in the text first and then specify where to put it, but this leaves no destination for it if the user exits other than by pressing NEXT. Ideally, perhaps, there would be a standard text editor associated with the exam system (and the comments system and the quiz system) so the user could move a cursor to anywhere in the answer area and enter answers.

P4. DO Loop and Array (author: B. Speelpenning)

There were two goals in the design of this PG/G, first to use interactive grading, and second to attempt to get maximum variability from a small generator. Both were met. On a few occasions students did get more points because the grading algorithm stopped them from propagating an error. On too many other occasions, however, students lost points from errors they would have spontaneously corrected and lost confidence through rejection of an answer. Maximum variability was gained, but at the expense of creating problems of greatly disparate difficulty.

a. Examples of disparate difficulty include that the predicate in the IF was never true in some versions, that the termination was sometimes by subscript-out-of-bounds, that the increment for M(I) was sometimes a constant and sometimes an array value, and that the array value increment was sometimes modified during a pass through the loop. The two versions used on the paper parts of the experiment illustrate the possible range of difficulty.

b. The answer entry mechanism required a specification that the last output had been generated (see proverb 53).

c. "del" was provided to delete an answer, but only applied to the first of the two values on a line. The subjects did not understand how to use this feature.

d. The interactive grading was a more severe problem on another exam when the interactive problem was not given as the last problem. Students tended to fret more about every answer on every problem after doing one problem interactively.

e. Two more problems with interactive grading were that people did correct themselves on the paper exam in ways that would have lost points on the interactive system, and that one subject entered the correct response to the second line while the system was waiting for a retry on the first line.

f. At least one subject lost points by pressing some key other than NEXT after entering an answer.

Appendix B. Questionnaires

Early in the experiment session subjects filled out the "Background Sheet" (p. 42). Their answers appear in Figure 4.1.

After the exams subjects filled out the "Questionnaire on PLATO Exam and on Paper Exam" (usually referred to as the post-test questionnaire) (pp. 43-46).

The "satisfaction" score in Figure 4.2 was derived from questions 1, 4, 5, 9, and 11-15 of the post-test. The weights were

question	response a	b	c	d	e
1	5	4	3	2	1
4	5	4	3	2	1
5	5	3	1		
9	1	3	5		
11	5	1	3	4	5
12	1	2	4	5	
13	1	2	3	4	5
14	5	1	3	3	
15	yes = 2			no = 4	

The average satisfaction was 3.11 with a standard deviation of .74.

Subject responses to post-test questions 1-15 appear in Figure 4.2. Their written comments in reply to question 16 are transcribed below, following the questionnaire.

Background sheet

For the evaluation of this experiment we need some biographical data about you.
We appreciate it very much if you answer the following questions completely.

All your answers will be kept confidential and used only for the evaluation of this experiment without mentioning your name.

1. What year were you born ? _____
2. What is your approximate college GPA ? _____
3. What year are you ?
   (Freshman, Soph., Jr., Sr., Grad.) _____
4. Which department are you in ? _____
5. How well are you doing in CS101 ?
   Poor _____
   fair _____
   good _____
   very good _____
6. What grade do you expect in CS101 ?
   (A, B, C, D, or E) _____
7. Approximately how many hours have you spent on PLATO ? _____
8. How well could you use a regular typewriter before using the PLATO keyboard ? (Check one, please.)
   a. never tried _____
   b. with great difficulty (spent a lot of time looking for keys) _____
   c. slow but sure (hunt-and-peck with only rare delays) _____
   d. good to expert (typed regularly) _____

Questionnaire on PLATO exam and on paper exam

1. How did you like the PLATO exam compared to the written exam ?

   a. liked the PLATO exam much more than the written exam
   b. " " " " a little more " " " "
   c. " " " " about the same as " " "
   d. " " " " a little less than " " "
   e. " " " " much less than the " "

2. What do you think about the environment during the exams?

   a. the experimental situation bothered me very much
   b. " " " bothered me sometimes
   c. " " " didn't bother me at all

3. What do you think of the content of the PLATO exam in general ?

   a. material tested was too difficult
   b. " " was challenging
   c. " " was of right difficulty
   d. " " was easy
   e. " " was too trivial

4. What do you think of the instructions and procedures for getting around and answering questions in the PLATO exam?

   a. very easy to follow
   b. easy to follow
   c. clear, but not obvious
   d. difficult to follow
   e. very confusing

5. Do you think you were able to concentrate on the PLATO exam ?

   a. I was able to concentrate
   b. PLATO disturbed my concentration sometimes
   c. I was not able to concentrate on the problems

6. What do you think of the content of the paper exam in general ?

   a. material tested was too difficult
   b. " " was challenging
   c. " " was of right difficulty
   d. " " was easy
   e. " " was too trivial

7. Do you think you were able to concentrate on the pencil and paper exam ?

   a. I was able to concentrate
   b. my concentration was disturbed sometimes
   c. I was not able to concentrate on the problems

8. How frequently did you switch to another problem on the PLATO exam ?

   a. very often
   b. sometimes
   c. only once to each problem

9. How much difficulty did you encounter when switching from one problem to another on the PLATO exam ?

   a. a lot of difficulty
   b. some difficulty
   c. very little difficulty

10. How frequently did you switch to another problem on the paper exam ?

    a. very often
    b. sometimes
    c. only once to each problem

11. Do you feel that the keyboard was an annoying hindrance in communicating with PLATO ?

    a. it was no hindrance for me
    b. it hindered me somewhat
    c. it was not more of a hindrance than a pencil
    d. it was easier to use than a pencil
    e. I like using the keyboard much more than using a pencil

12. Do you feel that your typing ability had an influence on your performance on the PLATO exam ?

    a. it had a great influence
    b. it had some influence
    c. it had almost no influence
    d. it had no influence at all

13. Do you feel that reading the problems from the screen took more time than reading them from paper ?

    a. it took much more time
    b. it took some more time
    c. I didn't feel a difference
    d. it took less time
    e. it took much less time

14. What kind of an exam would you prefer ?

    a. exam on PLATO
    b. paper and pencil exam
    c. part of exam on PLATO and part on paper
    d. I don't care

15. Was there anything on the PLATO exam that caused you more delay than you wanted ?

yes no

16. Please, write down anything that bothered you on the experiment or anything that bothered you on the exams or anything that caused you too much delay or any suggestions for further improvement of the exams.

We thank you very much for your help
and for your patience throughout the experiment

Responses to request for comments:

16. Please, write down anything that bothered you on the experiment or anything that bothered you on the exams or anything that caused you too much delay or any suggestions for further improvement of the exams.

SA: I feel that the directions were very confusing, especially with going back and changing answers and leaving a problem to go on to the next one without answering the previous problem. If these exams were ever used I feel that better instructions on how to use the machine would help alot.

SB: In the syntactical error part it was difficult to erase a wrong answer on the PLATO test.

SC: The examples used both in the Plato exercise and the written exam were relatively easy except for one section on Format which we haven't covered very much in class. Being conscious of being watched and recorded did make me apprehensive. I think this bothered me a little more on Plato because I'm not used to taking tests on it. I was less apprehensive with the paper and pencil test, I guess because I'm used to being watched while taking paper and pencil exams. Another aspect is with Plato you are "broadcasting" your answer to the world where in a paper and pencil situation if you think an instructor is trying to evaluate your answer while you write and this bothers you, you can cover it up easier.

SD: All of the instructions weren't clear. The atmosphere had some effect on concentration. I like the way in which you can return to problems. I think now that I have experienced the exam on Plato, I would do better on following ones. Being a totally different experience, it affected my output a little.

SE: Some instructions on Plato weren't very clear. I found an error in correction of syntax DO statement. Hadn't covered format material as of yet. Generally liked Plato exam though.

SF: I didn't mind the Plato exam. In fact, I think taking an exam on Plato could be quite enjoyable, but first, before I decide to take a real exam on Plato, I would have to gain a little more experience on Plato itself--its terminology, procedures, etc. This experiment seemed be, though, a bit "unfair" to Plato. We took an unfamiliar exam 1st on Plato and then on paper (after I knew what was going on). Maybe the exam should be "split" to where 1/2 of the exam is taken on Plato 1st and the other half on paper 1st.

SG: Syntax question difficult to work through due to problems with replace mode.

Appendix C. Experiment Administration

For each experimental session an "Activities" (p. 50) outline was drawn up prior to the experiment; the outline for the first session is included below. Subsequent sessions differed in the order exams were given and also in that only one tape was used for subsequent sessions.

The "User's Guide..." (p. 51) was given out because earlier experience had shown its necessity. In retrospect, the instructions should contain more details on how to perform the activities that are said to be possible.

The "Time Sheet" and "Log Sheet" (p. 52) were separate documents, but both served purely housekeeping functions.

Activities during experiment session I at
(Students:)
(Experimenters: Richard Doring, Larry Whitlock)

Experimenters	Students
1. Set up equipment (R)	
Hand out background sheet (L)	
Fill out time sheet (R)	Fill out background sheet
2. Hand out user's guide (L)	
Start practice exam (L)	
Record practice exam (R)	Take practice exam
Log practice exam (L)	
3. Hand out exam A_{Paper} (L)	
Record exam A_{Paper} (R)	Take exam A_{Paper}
Log exam A_{Paper} (L)	
Change & catalog tape (R)	
4. Start exam B_{PLATO} (L)	
Record exam B_{PLATO} (R)	Take exam B_{PLATO}
Log exam B_{PLATO} (L)	
Change & catalog tape (R)	
5. Hand out questionnaire (L)	
	Fill out questionnaire
Collect & catalog papers (11) (L)	

We have: 2 background sheets
2 time sheets
2 exams A_{Paper}
2 questionnaires
3 log sheets
11 papers + 2 tapes

User's guide for PLATO exams

These are a few hints about how you can proceed in the PLATO exam.

1) For all sections of the exam (other than the DO-loop problem), please remember :

- You can go back and forth to look at any problem or at any answer you gave.

- You can work at any problem at any time and as often as you wish.

- You can replace any answer you gave at any time during the whole exam.

- Your answers will not be judged or graded at once when you type them in.
  (Your answers will be graded only after you finish the whole exam .)

2) In the DO-loop problem only your answer will be judged immediately after you type it in .
A wrong answer can be corrected only once and you will get partial credit for that answer .

Time sheet

Name of student _____

Begin of experiment _____

Begin of practice exam _____
End of practice exam _____
Duration of practice exam _____

Begin of PLATO exam _____
End of PLATO exam _____
Duration of PLATOexam _____
Score

Begin of paper exam _____
End of paper exam _____
Duration of paper exam _____
Score

End of experiment _____

Log sheet

Session ____

Student _____

Content _____

Activities of the student: Time:

Appendix D. Definitions of Categorization of Activities

The following categories of activities were used to classify the subject behavior observed on the video tapes. Titles include brief name, code in parentheses, and long name.

Think (RT) Reading and Thinking: While the student reads the problem and thinks about the answers.

Calculate (CP) Calculating with Paper and Pencil: Whenever the student makes calculations on paper.

Answer (EA) Entering Answers: Includes time the student spent looking for the desired keys; deciding which key should be pressed next; waiting while answer is processed by the PG/G; and redoing answers the PG/G rejected (though some of this was recoded to Trouble).

Select (PS) Problem Selection: The period that begins with the keypress to leave a question and ends with the keypress to go to the next question. This period includes travel time to the cover page; time to display the cover page or instructions on how to get to the next question (e.g., in the syntax PG/G after the student pressed the key to indicate he was through working on the present question, a list of keys was displayed which permitted him to go to the next syntax question, the previous syntax question, or the syntax problem cover page); time to read the choices and make a selection; and time to enter the selection.

Most of the categories are well-defined, but it is appropriate to discuss the evidence we used to differentiate among Think, Answer, and Trouble. Usually during Think time the subject was motionless or fiddled with some object but stared at the exam or paper. Answer time was almost always time when the user bent the upper part of the body toward the screen/paper and moved the hand toward the answer mechanism. A perceptible relaxation of postural attitude usually marked the transition back to Think. Trouble time was identified by turning of the head toward the proctor and the asking of questions; it was also evident when the subject had completed an answer and was searching the screen and the surroundings for clues as to how to proceed. Knowledge of the various kinds of difficulties the subjects experienced--as recorded in the manual logs--also helped to identify Trouble time.

Appendix E. Detailed Logs of Subject Behavior

This appendix includes details of the behavior of each subject, although with far less detail for the unvideotaped subjects. For each taped subject there is a log, annotations for the log, and an "Analysis of Activities." Only the log annotations are available for the untaped subjects.

In deriving the logs presented here from the second-by-second analyses of the tapes, it was observed that occasionally Trouble time had been recorded as Answer entry time. To present a more appropriate picture of the amount of trouble the subjects had, a recoding rule was adopted: If an Answer time was inexplicably longer than ten seconds, it was reduced to six seconds and the rest of the time was counted as Trouble time (or sometimes Read and Think time). We further decided to count such converted intervals as separate activities which tended to slightly reduce the average size of some activity categories.

Each line of a detailed log corresponds to work on one problem. The time shown is the first second of starting to select the problem. (For confidentiality, all times have been shown as between eight and ten.) The problem numbers have been encoded with capital letters for PLATO and lower case for paper according to this code:

PLATO	paper	problem
P1	p1	arithmetic expressions
P2	p2	the entire syntax problem
Q1	q1	the first syntax subproblem
Q2	q2	the second syntax subproblem
Q3	q3	the third syntax subproblem
P3	p3	the PRINT/FORMAT problem
P4	p4	the interactive DO loop problem
quit		making the decision to terminate

The eight categories of activities fall naturally into the five basic categories shown in each log:

Select--There are three types: selection from the exam cover page, selection from the syntax problem cover page, and selection of the next syntax subproblem without going to the syntax cover page.

Display--Includes problem presentation, problem generation, and loading character sets. The latter two were small: Load was within a second of :19, Generation of P2 was about :29; Generation for P4 was only about :01.

Think & Calculate--Includes all time spent reading, thinking, and calculating.

Answer--Answer Entry time.

Trouble--Was encoded in the second-by-second logs as WN, "What next?".

Each entry in the log has three fields: time in minutes and seconds, number of activities, and notes. Only the first is usually present. The number of activities is preceded by a slash; if omitted, there was only one activity. Each note is preceded by an asterisk; the notes have these interpretation

*1 The time shown includes some problem Generation time.

*2 The time shown includes some character set Load time.

*3 The time shown does not include character set Load time because there was no load. Normally, at this point there would have been a Load, but the character set had been loaded during the practice exam.

*4 The time shown includes some time converted from another category as described above. This is usually mentioned in the annotations.

Detailed Log for Subject SA

PLATO exam

Wall Time	Pr #	Select	Display	Think & Calculate	Answer	Trouble
8:50:03	P1	:05	:02	2:05/7*4	:24/5	:16*4
:52:55	Q1	:22/2	:18/3*1*3	:18	:04	
:53:57	Q2	:01	:20/3*1	1:16/5	:14/3	2:18/6*4
:58:06	Q3	:02	:18/3*1	:06	:01	
:58:33	P3	:32/3	:08	:27		:13
:59:53	P4	:09	:08	5:34/11	:11/10	
9:05:55	P1	:15	:04	1:03		
:07:17	Q1	:16/2	:28/2*2	:07		
:08:08	Q2	:02	:07	:59/4	:19/4	2:41/5*4
:12:16	Q3	:02	:05	:07		
:12:30	P4	:20/2	:12	:05		
:13:07	quit	:07				

paper exam

Wall Time	Pr #	Select	Think & Calculate	Answer	Trouble
9:15:31	p1	:01	1:50/5	:07/5	
:17:29	q1	:01	:18/2	:02	
:17:50	q2	:01	:19		
:18:10	q3	:03	1:16/6	:07/3	
:19:36	p3	:01	:11		
:19:48	p4	:02	1:29/7	:12/6	
:21:31	q2	:06	:13	:01	
:21:51	q3	:01	:30/2	:05	
:22:27	p1	:07	:16		

Log Annotations for SA

practice exam

On the arithmetic expression problem, needed verbal explanation of why the arrow returned to the first question after entry to the last one. On the syntax problem, had trouble getting back to exam cover page. Spent 1:18 selecting problems and waiting for displays, 1:55 on productive work, and 2:17 on difficulties understanding instructions.

PLATO exam

8:50:03 P1 The 2:05 thinking time was composed mostly of calculation with paper and pencil (1:50). There was :07 pure reading and thinking time and :08 recoded from entering answer time. Illustrating a problem with the PLATO terminal, SA entered an underscore for a minus sign; this was quickly corrected though. (Underscore is a normal upper-case key; minus sign is on a black key at the left of the keyboard, together with plus, times, and divide.)

:53:57 Q2 Subject tried to give answer "E ASSIGN", meaning the word ASSIGN should be removed (it shouldn't, but subject didn't know the ASSIGN statement). PG/G would not accept that response and would not permit any action except to enter an acceptable "extra" character. Finally, we told the subject how to enter an acceptable character. Because BACK1 hadn't worked while the PG/G awaited the "extra" character, subject was then confused as to how to leave the PG/G. (The trouble time shown includes :39 recoded from entering-answer time.)

:59:53 P4 SA had considerable difficulty with the semantics of DO loops and the mechanics of interactive grading. After a :43 think, SA entered the wrong answer for the first line, but failed to notice that this

answer was rejected by the PG/G. After another :51 think, subject entered a response for second line, but PG/G assumed it was the retry for the first line and gave the correct answer to the first line. After thinking for another 1:07, the subject correctly entered the answer for the second line. After only :17, SA made an entry for the third line, but it was wrong because subject failed to observe that the array had been modified by the previous step. Noticing the rejection and taking great care, S2 re-entered an answer for line three after 1:23, but still got it wrong. (Answer expected was "sob" for subscript-out-of-bounds.) The two longest think times were recorded when subject was told previous response was wrong.

9:08:08 Q2 Subject attempted the "R"eplace option and had great difficulty; again because the PG/G would only accept certain characters as the characters to be deleted and inserted. Subject again had trouble getting out of the problem. (1:10 recoded from enter-Answer to Trouble.)

paper exam

9:15:31 p1 SA calculated all answers on paper. All the think time shown is calculate-with-pencil time. Only error was to get 27.0 as value of A/C/B*3.

:18:13 q3 The log record shows :08 calculate-with-pencil time, but no actual writing appears.

:19:36 p3 Did no work on this problem.

:19:48 p4 Each answer was written as it was computed. Subject got all answers except the value for M(5), which depended on the fact that the previous step had modified the value of M(3).

:21:31 q2 Answered correctly; had given no answer the first time through.

:21:51 q3 Changed answer to "M (9*31)".

:22:27 p1 The Select time was large because the subject flipped through many pages.

Analysis of Activities for SA

	Practice (first)			PLATO A (second)			Paper B (third)			Difference (PLATO-Paper)		
	a	b	c	a	b	c	a	b	c	a	b	c
Total	5:30	24	13.8	23:11	104	13.4	7:19	52	8.4	15:52	52	5.0
Reading and thinking	1:46	5	21.2	10:18	34	18.2	4:24	19	13.9	5:54	15	4.2
Calculating (pencil & paper)	0	-	-	1:49	5	21.8	1:58	7	16.9	-:09	-2	4.9
Entering answers	:09	3	3.0	1:13	24	3.0	0:34	17	2.0	:39	7	1.0
Total productive activities	1:55	8	14.4	13.20	63	12.7	6:56	43	9.7	6:24	20	3.0
Trouble time	2:09	4	32.3	5:28	13	25.2	0	-	-	5:28	13	25.2
Selection (exam cover page)	:27	3	9.0	1:20	8	10.0						
Selection (syntax cover page)	:17	2	8.5	:46	5	9.2				1:50	8	5.2
Other selection	:02	1	2.0	:07	4	1.8	:23	9	2.6			
Total organizational activities	2:55	10	17.5	7:41	24	19.2	:23	9	2.6	7:18	15	16.6
Loading characters	:19	1	19.0	:20	1	20.0						
Problem generation	:03	1	3.0	:33	3	11.0						
Problem display	:18	4	4.5	1:17	14	5.5						
Total PLATO activities	:40	6	6.7	2:10	18	7.2						
Scores				15/10/0/0⇒25/50			12/10/0/8⇒30/50					

a = Time (min.)
b = Number of Activities
c = Time per Activity (sec.)

Detailed Log for Subject SB

PLATO exam

Wall Time	Pr #	Select	Display	Think & Calculate	Answer	Trouble
8:45:01	P1	:09	:02	4:40/9	:26/6	
:50:18	Q1	:20/2	:10/2*1*3	:07/2*4	:06	
:51:01	Q2	:01	:21/2*1	:47	:01	
:52:11	Q3	:05	:14/2*1	1:13/7*4	:22/5	1:24/3*4
:55:29	Q1	:11	:08	:06		:34/2*4
:56:28	P3	:20/3*4	:04	1:26		
:58:18	P4	:12	:07	3:28/10	:48/9	:42/3*4
9:03:35	P3	:15	:08	:40		
:04:38	quit	:09				

paper exam

Wall Time	Pr #	Select	Think & Calculate	Answer	Trouble
8:27:49	p1	:01	3:51/7	:10/6	
:31:51	q1	:01	1:27/2	:01	
:33:20	q2	:01	2:50/6	:18/6	:08*4
:36:37	q3	:01	:37	:01	
:37:16	p3	:02	2:38/5	:54/5	
:40:50	p4	:01	1:39/6	:05/4	
:42:35	quit	:01			

Log Annotations for SB

practice exam

8:12:37 Spent 1:38/4 selecting problems, :32/4 waiting for display, 2:28/8 in reading and answering, and an enormous 5:46/3 trying to figure out what to do. 4:05 of the latter was because proctors did not yet understand that the syntax PG/G required an acceptable character after E. Subject had entered "E ." to indicate that the dots used to represent blanks were extra (i.e., subject did not understand the dot convention because had not used the compiler). Another 1:26 of Trouble time was spent trying to figure out how to get out of the syntax problem.

paper exam

8:27:49 p1 SB was fairly slow spending times of :25, 1:16, :18, :16, and 1:46; the last included :23 of calculation with pencil. In addition to pencil calculations, the subject used fingers to point to the problem and the variable definitions. The two long times included hard divisions: 18÷4.5 and 15÷4, respectively. Subject's only errors were to omit the sign from the third and fifth answers.

:31:51 q1 There was no error and the text was a simple assignment. Subject may have spent 1:28 thinking about it either because of worry that there should be an error or because unsure how to answer.

:33:20 q2 Problem was "ASSIGN 25 TO WUXYZIJH". Subject was very unsure, made several changes, and finally entered "E JH", which is correct but would have been rejected by the PG/G (in hand grading this was given one point). Subject also erroneously entered "R ASSIGN TO" so clearly understood neither the ASSIGN statement nor the R option on the PG/G. (:08 was recoded from entering Answer to Trouble time. Probably more should be moved from reading time to trouble time.)

8:38:31 p3 The five enter-Answer instances were long, totaling :54, so probably some time should be credited to reading or trouble. Answers were correct except for failure to round up on each of the three items.

:40:50 p4 Subject made no errors though did not mark that there were no more answers as would have been required on PLATO.

PLATO exam

8:45:01 P1 Subject did not point at problem and variables. One expression reduced to 2^8 and subject spent 1:04 working it out, including :10 calculating with pencil. There was no other pencil work; other expressions took reasonable times except the first took 1:19. Subject changed one answer.

:52:11 Q3 "DO 25 KD = 93 13" Here the syntax PG/G has made an error: this is a legal assignment or else it should be a DO statement with limits "9,313" since the upper limit must exceed the lower. SB tried "E D" apparently thinking that control variables had to be one letter. (This points up a teaching deficiency.) Because the system rejected that first answer, the subject reflected until coming up with a correct answer, "M ,"; thus, subject got credit on PLATO where none would have been given on paper. (Some time originally coded as enter-Answer has been recoded as :10 reading-Thinking and 1:00 Trouble.)

:55:29 Q1 Subject returned to Q1 by mistake while trying to leave the syntax problem. Thus, all this time--an additional :25--could be counted as trouble time. (Indeed, :28 of the trouble time shown has been recoded from the selection time for the next problem.)

:56:28 P3 Subject considered this problem for some time, but made no answers. Apparently didn't understand I format items, even though understood F items on the paper exam.

8:58:18 P4 Subject entered first answer, but neglected to press NEXT. (:42 of the time spent waiting has been recoded from enter-Answer to Trouble time.) Subject got first answer wrong twice so system gave the answer. Log claims second response was correct, but subject got no points; presumably because they were all subtracted elsewhere. Later SB typed "sob" (correct) but hit shift-BACK to return from PG/G so answer was not recorded by system.

post-test questionnaire

Responding to question 11 as to whether the keyboard was an "annoying hindrance," the subject circled a response code for "somewhat." We asked verbally what the problem was and the subject mentioned only the trouble with the syntax problem; i.e., the "hindrance" was in the problem design rather than the keyboard.

Verbal questioning about typing ability brought out the fact that the subject had had a typing class three years previous.

In other verbal comments the subject mentioned having learned F FORMATs from reading, but not having learned I FORMATs (which explains the better performance on p3 even though it was harder than P3).

Analysis of Activities for SB

	Practice (first)			PLATO B (third)			Paper A (second)			Difference (PLATO-Paper)		
	a	b	c	a	b	c	a	b	c	a	b	c
Total	10:24	20	31.2	19:46	85	14.0	14:47	58	15.3	4:59	27	-1.3
Reading and thinking	1:53	4	28.3	12:17	31	23.8	12.39	26	29.2	-:22	5	-5.4
Calculating (pencil & paper)	0	-	-	:10	1	10.0	:23	1	23.0	-:13	0	-13.0
Entering answers	:35	4	8.8	1:43	22	4.7	1:29	23	3.9	:14	-1	.8
Total productive activities	2:28	8	18.5	14:10	54	15.7	14:31	50	17.4	-:21	4	-1.7
Trouble time	5:48	3	116.0	2:40	8	20.0	:08	1	8.0	2:32	7	12.0
Selection (exam cover page)	1:04	3	21.3	1:10	6	11.7						
Selection (syntax cover page)	:34	2	17.0	:20	2	10.0				1:34	5	7.4
Other selection	0	-	-	:12	4	3.0	:08	7	1.1			
Total organizational activities	7:26	8	55.8	4:22	20	13.1	:16	8	2.0	4:06	12	11.1
Loading characters	:20	1	20.0	0	-	-						
Problem generation	:07	1	7.0	:31	3	10.3						
Problem display	:05	2	2.5	:43	8	5.4						
Total PLATO activities	:32	4	8.0	1:14	11	6.7						
Scores				9/15/0/0⇒24/50			9/11/5/10⇒35/50					

a = Time (min.)
b = Number of Activities
c = Time per Activity (sec.)

Detailed Log for Subject SC

PLATO exam

Wall Time	Pr #	Select	Display	Think & Calculate	Answer	Trouble
8:44:01	P1	:02	:03	4:19/14	:12/6	:05
:48:42	Q1	:22/2	:12/3*1*3	:42	:01	
:49:59	Q2	:03	:17/3*1	1:10/3	:05/3	1:25
:52:59	Q3	:02	:16/3*1	1:30/3	:03/2	:11
:55:01	P3	:28/3	:07	1:35		:03
:57:14	P4	:05	:10/2*1	5:16/11	:15/9	:52/2
9:03:52	P1	:17	:04	1:27		
:05:40	Q1	:14/2	:24/2*2	:39		
:06:57	Q2	:08	:06	:23		
:07:34	Q3	:02	:08	:29		
:08:13	Q1	:05	:08			
:08:26	P4	:29/2	:10	:05		
:09:10	P3	:04	:13	1:08		
:10:35	quit	:10				

paper exam

Wall Time	Pr #	Select	Think & Calculate	Answer	Trouble
9:12:40	p1	:01	3:23/11	:08/5	
:16:12	q1	:01	:44	:02	
:16:59	q2	:02	1:28/6	:26/4	
:18:55	q3	:02	:53/3	:03/2	
:19:53	p3	:02	:06		
:20:01	p4	:02	1:43/5	:08/4	
:21:54	p3	:01	:02		
:21:57	q3	:01	:09		
:22:07	q2	:01	:13		
:22:21	q1	:01	:07		
:22:29	p1	:01	:53/5	:02	
:23:25	quit	:02			

Log Annotations for SC

practice exam

Subject spent 1:44 selecting and waiting for problems, 2:01 doing work, and 3:38 wondering and asking what to do.

P1 Needed verbal explanation of how to use BACK1 to leave a problem.

P2 Also needed explanation of the difference between dots (meaning spaces) and periods; also how to erase an answer.

Quit once but wondered what to do and went back to cover page for :40.

PLATO exam

8:44:01 P1 SC simply seemed slow at arithmetic; spent large amount of time with pencil calculations (:51); also spent long time (:53) checking after making last answer entry.

:49:59 Q2 Problem was "SI = 23**PO + MFP" SC had entered "E" (for Extra "+ MFP") and changed mind. The syntax PG/G gave the usual trouble by demanding that a character be specified.

:52:59 Q3 Spent a long time checking after entering answer. Seemed to look for another error; possibly because the arrow moved down.

:57:14 P4 DO loop PG/G rejected SC's first attempt at the first answer. This rejection gave SC considerable pause. Spent at least a minute worrying about it. After asking proctor and being told to try again, got it wrong again. Subsequent lines were entered correctly (including "sob"), but grading algorithm had already reduced score to zero.

9:08:13 Q1 Subject had wanted to get out of P2. NEXT1 went to Q1 rather than P3.

paper exam

9:12:40 p1 Did calculations (1:55) on separate page and on exam. Only error was wrong sign on third problem.

:16:59 q2 "ASSIGN 25 TO WUZXIYHJ" The correct response "Extra HJ" would not have been accepted by the system. SC entered this response okay on paper, but also wrote "R =". In addition, SC wrote a comment (the long enter-answer time (:26)): "Also variable name should be first WUZXIY = 25".

:18:55 q3 Gave incorrect answer: "E space" and drew line under space in "PRINT ,".

:20:01 p4 Gave three lines of output, failed to comprehend the IF...GOTO.

Analysis of Activities for SC

	Practice (first)			PLATO B (second)			Paper A (third)			Difference (PLATO-Paper)		
	a	b	c	a	b	c	a	b	c	a	b	c
Total	7:23	30	14.8	26:44	106	15.1	10:47	65	9.95	15:57	41	5.15
Reading and thinking	1:45	6	17.5	17:52	36	29.8	7:24	29	15.3	10:28	7	14.5
Calculating (pencil & paper)	0	-	-	:51	3	17.0	2:17	7	19.6	-1:26	-4	-2.6
Entering answers	:22	5	4.4	:36	21	1.7	:49	17	2.9	-:13	4	-1.2
Total productive activities	2:07	11	11.5	19:19	60	19.3	10:30	53	11.9	8:49	7	7.4
Trouble time	3:20	7	28.6	2:36	6	26.0	0	-	-	2:36	6	26.0
Selection (exam cover page)	:52	4	13.0	1:27	9	9.7						
Selection (syntax cover page)	:21	2	10.5	:39	4	9.8				2:14	7	6.5
Other selection	:08	1	8.0	:25	6	4.2	:17	12	1.4			
Total organizational activities	4:41	14	20.1	5:07	25	12.3	:17	12	1.4	4:50	13	10.9
Loading characters	:23	1	23.0	:19	1	19.0						
Problem generation	:05	1	5.0	:26	4	6.5						
Problem display	:07	3	2.3	1:33	16	5.8						
Total PLATO activities	:35	5	7.0	2:18	21	6.6						
Scores				12/15/0/0⇒27/50			12/6/0/7⇒25/50					

a = Time (min.)
b = Number of Activities
c = Time per Activity (sec.)

Detailed Log for Subject SD

PLATO exam

Wall Time	Pr #	Select	Display	Think & Calculate	Answer	Trouble
8:36:45	P1	:13	:09	2:01/8	:24/6	
:39:32	Q1	:22/2	:19/3*1*3	:15	:01	
:40:29	Q2	:03	:13/3*1	:35	:01	
:41:21	Q3	:08	:13/3*1	:14	:01	
:41:57	Q1	:01	:08			:14
:42:20	P3	:24/3	:08	1:50/10	:37/8	:28
:45:47	P4	:11	:10/2*1	1:45/6	:14/5	:22
:48:29	P1	:13	:02	1:43/5	:05/3	
:50:32	Q1	:16/2	:24/2*2	:22		
:51:34	Q2	:05/2	:09	:05		
:51:53	Q3	:07/2	:08	:43/4	:04/3	
:52:55	P3	:22/3	:08	:35		
:54:00	P4	:08	:09	:11		
:54:28	quit	:09				

paper exam

Wall Time	Pr #	Select	Think & Calculate	Answer	Trouble
8:24:17	p1	:02	3:09/2	:09	:04
:27:41	q1	:02	:25/2	:02	:12
:28:22	q2	:03	:25	:02	
:28:52	q3	:04	1:00/4	:18/4	:28
:30:42	p3	:04	1:05/4	:21/3	
:32:12	p4	:03	2:51/9	:27/9	
:35:33	quit	:05			

Log Annotations for SD

practice exam

8:18 Q2 Had considerable trouble trying to return from syntax problem to exam cover page (shift-NEXT would return to the same problem because there was only one). Spent 1:47 in productive work and 2:03 trying to figure out what to do next.

paper exam

8:24:17 p1 Spent 2:46 calculating all answers with pencil (not on exam paper though) then :04 inquiring how to enter answer and :09 entering all answers at once. Made three mistakes, all by wrong sign; two were wrong sign on final answer and one was wrong sign for intermediate result.

:27:41 q1 The "trouble" time here was a question as to whether the syntax errors might simply be that the text was in the wrong columns (they were not). In fact, q1 and q2 had no syntax errors.

:29:50 q3 The :28 trouble time was uncertainty as to how to fill out a "replace" error response. Eventually did correctly. Next subject entered "Missing limit," but correctly erased that.

:32:12 p4 Of the nine answer entries, seven took one second and two took ten. SD started to respond with values of 1, 2, 3,..., but then spontaneously noticed the increment field. Had SD been interactively graded, the error message for the second row might have caused more worry and concern than enlightenment. Got the third line wrong because an array value it used was changed by calculation for second line. Failed to specify "sob".

PLATO exam

8:36:45 P1 Did :32 calculation with paper for first expression, but did the rest in head. (May have worked on more than one problem while calculating on paper.)

:41:57 Q1 The second visit to Q1 was unwanted; a press of NEXT from the syntax cover page takes one to it. This could be recorded as an additional :09 of Trouble time.

:44:02 P3 SD required verbal explanation of how to enter lines by first entering the line number where they were to go.

:47:06 P4 Pressed shift-BACK when should have pressed NEXT to terminate the answer and have it accepted. Subject recovered, entered the answer correctly, and got full credit.

:49:02 P1 Again SD calculated answers with pencil (:23); also changed some answers.

:52:08 Q3 Changed answer.

No change to P3 or P4 during second visit.

Analysis of Activities for SD

	Practice (first)			PLATO A (third)			Paper B (second)			Difference (PLATO-Paper)		
	a	b	c	a	b	c	a	b	c	a	b	c
Total	5:34	27	12.4	17:52	113	9.5	11:21	51	13.4	6:31	62	-3.9
Reading and thinking	1:37	5	19.4	9:24	37	15.2	6:09	21	17.6	3:15	16	-2.4
Calculating (pencil & paper)	0	-	-	:55	2	27.5	2:46	1	166.0	-1:51	1	-138.5
Entering answers	:10	3	3.3	1:27	28	3.1	1:19	19	4.2	:08	9	-1.1
Total productive activities	1:47	8	13.4	11:46	67	10.5	10:14	41	15.0	1:32	26	-4.5
Trouble time	2:03	5	24.6	1:04	3	21.3	:44	3	14.7	:20	0	6.6
Selection (exam cover page)	:12	2	6.0	1:32	9	10.2						
Selection (syntax cover page)	:33	3	11.0	:47	6	7.8				2:19	15	4.1
Other selection	:11	2	5.5	:23	7	3.3	:23	7	3.3			
Total organizational activities	2:59	12	14.9	3.46	25	9.0	1:07	10	6.7	2:39	15	2.3
Loading characters	:18	1	18.0	:18	1	18.0						
Problem generation	:05	1	5.0	:28	4	7.0						
Problem display	:25	5	5.0	1:34	16	5.9						
Total PLATO activities	:48	7	6.9	2:20	21	6.7						
Scores				9/10/9/10⇒38/50			6/15/10/8⇒39/50					

a = Time (min.)
b = Number of Activities
c = Time per Activity (sec.)

Annotated Log for Subject SE

paper exam (8:27-8:41)

p1 Although subject did paper calculation on second subproblem, the answer given was inexplicably 17.7 instead of the correct 1.0. The answer -3.0 for the third subproblem might possibly be by confusion between the letter I and the digit one. For the fourth subproblem SE answered "27" where the answer should be REAL and 3; this is probably due to wrong order for double exponentiation. The incorrect -.75 given for the fifth subproblem was probably due to missing the sign on the operand.

q2 Gave the answer "E H J" which is the best possible, but we graded it zero as would the PG/G. (Had it been a video-taped exam, we would have given it one point when we went through and regraded this problem for further analysis.)

p3 Without comment, we present the subject's response:

```
1     F5.2   F4.1   F3.0
      7.3862 7.85   8.74
```

p4 Gave answers for 1, 3, and 5, but later went back and corrected it by erasing the middle line (it is skipped by the IF...GOTO). Also changed first line from 1, 6 to the correct values of 1, 2. Subject would have received no credit from the interactive grading scheme.

PLATO exam (8:42-9:06)

8:59 P2 Subject had trouble changing the answer to the syntax problem.

9:04 P4 Tried to correct an answer (ERASE is permitted until the line is finished), but found the instructions confusing.

9:06 Spent ten seconds staring at the page intermediate between leaving the cover page and actually quitting the exam. This was either spent trying to understand that yet another keypress was needed to really leave the exam or trying to review the problems on the exam.

post-test questionnaire

12 Verbally the subject reported having typed a number of papers.

13 SE reported not feeling a difference in reading problems from the screen; and yet spent ten minutes longer on PLATO. Subject further mentioned that there was trouble switching problems, which could account for the longer time.

14 Subject did prefer exams on PLATO because they are "more interesting," and "more exciting."

Annotated Log for Subject SF

We took a fairly extensive log of SF's PLATO work and will present much of it here. Even though it does not reveal any particular anomolies, it does illustrate the level of analysis that can be conducted without videotape.

practice exam (8:36-8:44)

8:36 Cover page

:36:30 P1

:40:25 Cover page

:40:35 Selects syntax cover page. Character loading begins.

:40:50 Syntax cover page

:41:30 FORTRAN syntax directions (few students took this option)

:42:00 Syntax cover page

:42:45 Q1

:44:00 Syntax cover

:44:05 Cover page

PLATO exam (8:46-9:02)

8:46:10 P1 Thought for awhile; gave answers very fast.

:49:55 Cover page

:50:05 Syntax cover page

:50:08 Q1 (no error in problem)

:51:00 Q2 (no error in problem)

:52:00 Q3 (there were errors, but SF found them)

:53:05 Syntax cover page

8:53:20 Q1 Review (The log says "review," but the subject was probably just having trouble getting to P3.)

:53:25 Syntax cover page

:53:40 Cover page

:53:48 P3 Studied page for two minutes before typing.

:56:15 P3 Finished

:56:28 Cover page (Presumably it took thirteen seconds to display.)

:56:40 P4

:57:35 P4 Entered first answer.

:58:15 P4 Entered second answer.

:58:30 P4 Made an incorrect entry for third answer.

9:01:05 P4 Entered third answer correctly.

:01:15 Cover page

:01:30 Quit

:01:40 Quit (Pressed last exit button.)

paper exam (9:08-9:18)

p1 Lost three points by incorrectly copying correct answer from work area to answer area. (Gave answer as -3.15.)

q2 Gave answer "E H J" and received zero credit.

p3 Failed to round twice and used wrong format for third answer. (Had "8.7")

p4 Wrote correct answers but did not indicate the end of output as would have been required on the PG/G for full credit. Got full credit anyway.

Annotated Log for Subject SG

practice exam (8:15-8:26)

Did the two arithmetic expressions in two minutes, but spent eight minutes looking for syntax errors in "81....CONTINUE".

paper exam (8:28-8:49)

p1 The only calculation written on the exam is "-15./4." for the third problem. All answers correct.

p2, p3 All answers correct; no other marks on paper.

p4 Modified the values in the array as the computation proceeded. Got all answers correct (even though one iteration modified an array value used by the next).

PLATO exam (8:51-9:06)

8:51 Begin at monitor

:51:30 Cover page

:52 P1 Motionless while thinking of answers. Typed answers very fast.

:54:30 Corrected wrong answer. Appeared to be disturbed by the sound of the machine (probably the videotape machine and the nature of observational situation).

:55:15 Cover page

:55:35 Q1 No syntax error; answered very fast.

:58-:59 Q2-Q3 Was observably confused by the instructions. Did do Q3 in only 0:15.

8:59 P2 Reviewed answers

9:00:05 Syntax cover page

:00:15 Cover page

:00:20 P3 Was confused by the replace-a-line scheme.

:02 Cover page

:02:30 P4 Worked with paper and pencil some. Answer entries were at :03:55, :03:58, :04:20, and :04:30.

:04:40 Cover page

:05:00 Review

:05:35 Detected and corrected an error.

:05:40 Finished

post-test questionnaire

12 Wrote on the questionnaire that typing ability influenced performance by "increased speed + fewer clerical errors only".

BIBLIOGRAPHIC DATA SHEET	1. Report No. UIUCDCS-R-76-782	2.	3. Recipient's Accession No.
4. Title and Subtitle DETAILS OF AN EXPERIMENTAL VIDEOTAPE EVALUATION OF AN INTERACTIVE EXAM SYSTEM			5. Report Date December 1976 6.
7. Author(s) Richard Doring, Lawrence R. Whitlock, Wilfred J. Hansen			8. Performing Organization Rept. No. UIUCDCS-R-76-782
9. Performing Organization Name and Address National Science Foundation Washington, DC			10. Project/Task/Work Unit No. 11. Contract/Grant No. EC41511
12. Sponsoring Organization Name and Address Department of Computer Science University of Illinois Urbana, Illinois 61801			13. Type of Report & Period Covered Research 14.

15. Supplementary Notes

16. Abstracts

This report provides details of an experiment conducted at the University of Illinois. An analysis of the data is contained in a companion paper (UIUCDCS-R-76-836) about the student performance on an interactive examination.

17. Key Words and Document Analysis. 17a. Descriptors

PLATO system
automated computer science education system

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
unlimited	UNCLASSIFIED	85
	20. Security Class (This Page) UNCLASSIFIED	22. Price

OCT 1 7 REC'D

BOUND BY HNM

UNIVERSITY OF ILLINOIS-URBANA
510.84 IL6R no. C002 no.782-787(1976
Design of Irredundant MOS Networks : a p
3 0112 088402588